

PMC-VQA: Visual Instruction Tuning for Medical Visual Question Answering

Year: 2023 | Citations: 256 | Authors: Xiaoman Zhang, Chaoyi Wu, Ziheng Zhao, Weixiong Lin, Ya Zhang

Abstract

Medical Visual Question Answering (MedVQA) presents a significant opportunity to enhance diagnostic accuracy and healthcare delivery by leveraging artificial intelligence to interpret and answer questions based on medical images. In this study, we reframe the problem of MedVQA as a generation task that naturally follows the human-machine interaction and propose a generative-based model for medical visual understanding by aligning visual information from a pre-trained vision encoder with a large language model. We establish a scalable pipeline to construct a large-scale medical visual question-answering dataset, named PMC-VQA, which contains 227k VQA pairs of 149k images that cover various modalities or diseases. We train the proposed model on PMC-VQA and then fine-tune it on multiple public benchmarks, e.g., VQA-RAD, SLAKE, and Image-Clef-2019, significantly outperforming existing MedVQA models in generating relevant, accurate free-form answers. In addition, we propose a test set that has undergone manual verification, which is significantly more challenging, serving to better monitor the development of generative MedVQA methods. To facilitate comprehensive evaluation and comparison, we have maintained a leaderboard at <https://paperswithcode.com/paper/pmc-vqa-visual-instruction-tuning-for-medical>, offering a centralized resource for tracking progress and benchmarking state-of-the-art approaches. The PMC-VQA dataset emerges as a vital resource for the field of research, and the MedVInT presents a significant breakthrough in the area of MedVQA.